CS 134: Networks — Problem set 3

Prof. Yaron Singer — Due: **Wednesday 2/22/2017**

1. Non-Preferential Attachment (20 points). Consider the following model of non-preferential attachment networks:

- At time $t = 1$, there is one node (node 1) in the network.
- At time $t > 1$, node t is added and it establishes one directed edge from node t to $i_1, \cdots, i_{t-1}$ where each i_j is picked independently and uniformly at random between 1 and $t - 1$.

a. (8 points) At time step t, what is the expected in-degree[1] of node $i_k, k \in \{1, \ldots, t-1\}$?

b. (8 points) At time step t, what is the fraction of nodes that have expected in-degree at least d?

c. (4 points) Do you expect graphs generated using this model to have a power-law degree distribution? Briefly explain why or why not.

Note: In this problem, your answers should be closed-form formulas (e.g., they should not contain a summation from 1 to t). You may treat the following approximation as if it were exact: $\sum_{i=1}^{n} \frac{1}{i} \approx \log(n)$.

2. Power Laws in Classrooms [Easley and Kleinberg, 18.8 Q3] (10 points). Suppose that some researchers studying educational institutions decide to collect data to address the following two questions.

- As a function of k, what fraction of Harvard classes have k students enrolled?
- As a function of k, what fraction of third-grade elementary school classrooms in New York State have k pupils?

Which one of these would you expect to more closely follow a power-law distribution as a function of k? Give a brief explanation for your answer, using some of the ideas about power-law distributions.

3. It's a Power-Law Story (15 points). Suppose for \$100 you can purchase the exclusive rights to sell the latest single by Taylor Swift. However, due to branding restrictions you can only sell the single for \$1 to each buyer. You are now trying to decide whether or not purchasing the rights for \$100 would be financially wise. Of course, you don't know how many people will actually purchase the song, which we can model as a random variable X. After doing some research, you discover that the likelihood of a single having exactly k purchases follows a power-law distribution with parameters $c = 1$ and $\alpha = 2$ for all values of $k \in \{1, \cdots, 485165195\}$, with the likelihood of any more purchases being zero (you read a recent survey in the Economist tipping you off that there

[1] The in-degree of a node u is the number of edges directed towards u.

are exactly 485,165,195 people in the world who are willing to pay for music; the rest follow their conscience and try to get it for free). Assume that your expected profit is your expected revenue from selling the single minus the cost for the sales rights (in this case, \$100). If you're trying to make a positive profit, should you purchase the rights?

4. Power-Law Cliques (25 points). Fix a graph $G = (V, E)$. Construct a graph $\widetilde{G}$ as follows. For every node $i \in V$ with degree k, let $\widetilde{G}$ contain a clique C_i of $k + 1$ nodes. These cliques are all distinct, and $\widetilde{G}$ is the union of all of them.

a. (3 points) Suppose $V = \{a, b, c, d\}$ and $E = \{ab, bc, ac, ad\}$. Draw $\widetilde{G}$.

b. (7 points) For an arbitrary G, let p be the degree distribution of G, so that $p(k)$ is the fraction of nodes with degree k in G. Denote by Δ the maximum degree in G. Let $\widetilde{p}$ denote the degree distribution of $\widetilde{G}$. Give a formula for $\widetilde{p}(k)$ in terms of p (i.e., in terms of the values of p).

c. (8 points) We say a distribution q on the natural numbers has the *power-law property until* K if there is a constant $\beta > 1$ so that

$$\frac{q(k')}{q(k)} \geq \left(\frac{k'}{k}\right)^{-\beta}$$

whenever k and k' are integers such that $1 \leq k \leq k' \leq K$. You can check that this is consistent with the definition of a power-law distribution from lecture.

Suppose that p has the power-law property until Δ. Show that $\widetilde{p}$ has the power-law property until Δ as well (possibly with a different constant β).

d. (7 points) The "strong form" of the friendship paradox says that for a typical person i, the average degree of i's neighbors will strictly exceed i's own degree. Does this necessarily hold in networks whose degree distributions have the power law property?

5. Coding: Fitting Power-Laws (30 points). In this coding problem, we will use the files network1.txt and network2.txt in the pset folder. In each file, the two numbers per line represent an edge from the first node to the second. For all plotting, we advise you to use matplotlib's pyplot function.

a. (5 points) For each data set, plot the degree distribution of the social network with the degree d on the horizontal axis and the fraction of nodes with degree d on the vertical axis, i.e., $f(d) = \frac{|\{v \in V : \text{v has degree } d\}|}{|V|}$.

b. (5 points) For each data set, plot the degree distribution of the graph in log-log scale. That is, for d and $f(d)$ as above, plot $\log d$ on the horizontal axis and $\log f(d)$ on the vertical axis. Does the plot seem linear?

c. (7 points) Using your $\log - \log$ plot, fit the following simplified power-law model using ordinary least squares regression (see section notes for details) with an appropriately chosen bin size for : $\log p(x) = \alpha \log x + b$ where b is some constant.

d. (10 points) For each data set, fit the power-law model to the data sets using the method of maximum likelihood. Specifically, write a method which, given an array of "degrees" (where the ith entry of this array is the degree of node i), fits the parameter α of the power law model to the given input.

e. (3 points) Plot the power law distributions for your estimates for parameters from parts (c) and (d). Briefly discuss the qualitative differences between the two methods, why these difference might happen, and the appropriateness of a power law model for these datasets.